



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/607,370	06/30/2000	Reiner Kraft	ARC9-2000-0046-US1	8705
23334 7590 11/29/2007 FLEIT, KAIN, GIBBONS, GUTMAN, BONGINI & BIANCO P.L. ONE BOCA COMMERCE CENTER 551 NORTHWEST 77TH STREET, SUITE 111 BOCA RATON, FL 33487			EXAMINER STORK, KYLE R	
			ART UNIT 2178	PAPER NUMBER
			NOTIFICATION DATE 11/29/2007	DELIVERY MODE ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

ptoboca@focusonip.com

Office Action Summary

Application No.

09/607,370

Applicant(s)

KRAFT ET AL.

Examiner

Kyle R. Stork

Art Unit

2178

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 22 October 2007.
- 2a) ☐ This action is FINAL. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 3-13, 16-19 and 21-25 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 3-13, 16-19, and 21-25 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

1. This non-final office action is in response to the Request for Continued Examination filed 22 October 2007.
2. Claims 3-13, 16-19, and 21-25 are pending. Claims 21-25 are newly added. Claims 21, 23, and 25 are independent claims.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 3, 7-11, 13, 16, 21, 23, 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Meyerzon et al. (herein after Meyerzon) U.S. Patent No. 6,638,314 B1 filed 6/26/1998 in further view of Blumenthal U.S. Patent No. 6,026,409 filed 9/26/1996 and further in view of Koike et al. (US 7194678, filed 1 March 2000, hereafter Koike)

As per independent claim 21, Meyerzon discloses a method for indexing data documents, the method comprising:

Retrieving, to a server, with a web crawler from a network address, a data document with client-side scripting code therein (Figure 2: Here, a web crawler server is implemented between a client and a web server)

Executing, at the server, a web-browser, as part of the web crawler, wherein the web-browser displays an in-memory copy of the data document which has been retrieved, wherein the in-memory copy of the data document maintains a web-browser display format and a web-browser display layout of the dynamic data document when displayed in the web browser (Meyerzon Col 7 Lines 60-65 and Col 8 Lines 15-20: Here, the crawler acts as a web browser in that it requests the web page data. These requested web page documents are stored in memory in a display format)

Executing, at the server instead of a client system, a browser scripting engine as part of the web-browser for loading content as directed by the client-side scripting code into the in-memory copy creating a final web-browser display representation of the dynamic data document so that the final web-browser display representation is substantially similar to when the data document is viewed by a user in the user's web-browser running on the client system when all the data is viewed (Meyerzon Col 7 Lines 60-65 and Col 8 Lines 15-20)

Indexing, at the server, the content in the memory, wherein the content being indexed is the content which has been loaded by the browser scripting engine in order to index the data document as if being viewed by the user in the user's web-browser on the client system (Figures 4-5).

Meyerzon does not specifically mention *wherein the server processing unit renders the in-memory webpage prior to analyzing and summarizing the in-memory webpage*. However, Blumenthal mentions a document that can be rendered prior to user actions (Blumenthal Col 17 Lines 45-53). It would have been obvious to one of

ordinary skill in the art at the time of the invention, to apply Blumenthal to Meyerzon, providing Meyerzon the benefit of rendering the document prior to the user action to ensure the correct page is being analyzed and summarized.

Meyerzon does not specifically disclose wherein the data document is a dynamic data document. However, Koike discloses a proxy server assembling a dynamic data document for display at a client browser (Figure 6; column 7, lines 13-33). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to have combined Koike with Meyerzon, since it would have allowed a user to more quickly receive the dynamic data.

In regard to dependent claim 3, Meyerzon further discloses *wherein the one or more images with textual content embedded therein include at least one of an in-line GIF image and an in-line JPEG image*. (Meyerzon Col 9 Lines 37-46 i.e. an image is retrieved to display on a web page and it is well known in the art the images displayed on web pages can be a gif and jpeg image).

In regard to dependent claim 7, Meyerzon further discloses *initializing a first list with seed values* (Meyerzon Col 17 Lines 25-26 i.e. assigning a current crawl number to the current web crawl); *checking if there are any URLs to be processed and in response that any URL exists to be processed then performing the follow sub-steps of* (Meyerzon Col 17 Lines 28-29 i.e. determine whether an electronic document has been retrieved): *determining if a URL is in a second list; and in response that a URL is not in the second list then performing the following sub-steps of: inserting the URL into the first list; scheduling the URL for crawling; crawling the URL when scheduled to do so; removing*

Art Unit: 2178

the URL from the first list after the scheduled crawling; entering the URL into the second list (Meyerzon Col 9 Lines 64 and Col 10 Lines 1-11 i.e. history map checks each hyperlink URL to determine if it is already listed in the history map, if not the URLs are added and are marked as not being crawled and added to the transaction log. The history map includes a number crawled and number modified data); and repeating the checking step until there are no more URLs to be processed; where if the determining step determines that the URL is in the second list then repeating the checking step until there are no more URLs to be processed. (Meyerzon Col 12 Lines 1-17 i.e. retrieves and processed a URL until there are none left in the transaction log)

In regard to dependent claim 8, Meyerzon further discloses *wherein the sub-step of initializing a first list with seed values further includes the list being a URL pool.* (Meyerzon Col 7 Lines 65-67 i.e. retrieving a processing URLs from the transaction log)

In regard to dependent claim 9, Meyerzon further discloses *wherein the sub-step of determining if a URL is in a second list further includes the second list being a visited pool.* (Meyerzon Figure 4 shows a column indicating the number crawled and modified)

In regard to dependent claim 10, Meyerzon discloses *wherein the sub-step of crawling further comprises the sub-steps of: issuing an HTTP command to a web server named in the URL; receiving contents of an HTML page as a result of the issued HTTP command; and passing on the contents of the HTML page to a Page Rendering subroutine.* (Meyerzon Col 8 Lines 26-35 i.e. the client computer transmits data to a search engine, the search engine examines its associated index to find documents and

Art Unit: 2178

returns the documents which are secondary documents and lists the documents for the user to view)

In regard to dependent claim 11, Meyerzon discloses *receiving the contents of the HTML page in the Page Rendering subroutine; building an in-memory representation of a layout for the HTML page and if more data is needed to properly form the representation, then performing the sub-steps of* (Meyerzon Col 7 Lines 60-65 and Col 8 Lines 15-20 i.e. web crawler program searches remote server computers connected to the network for electronic documents and retrieves electronic documents and associated data and a browser displays documents to a user): *requesting additional web-based information; gathering this additional web-based information; inserting any URLs associated with this additional web-based information into the second list and a URL cache* (Meyerzon Col 9 Lines 37-46 i.e. an image is retrieved to display on a web page); *building a final amended representation; and forwarding the final amended representation to an Extraction subroutine; wherein, if no more data is needed to properly form the in-memory representation, then forwarding the in-memory representation to the Extraction subroutine.* (Meyerzon Col 16 Lines 32-44)

In regard to dependent claim 13, Meyerzon discloses *receiving a text map from the Page Extractor subroutine; processing the text map in an application-specific manner* (Meyerzon Col 2 Lines 48-51 i.e. information from the electronic document retrieved from the web crawl is stored in an index to begin the routine); *applying data extraction patterns to the text map* (Meyerzon Col 5 Lines 7-8 i.e. extracting data from each of the retrieved documents); *translating resultant data from the applying step;*

Art Unit: 2178

forwarding any URLs present in the text map to a manager subroutine; and forwarding any extracted data and metadata to application logic. (Meyerzon Col 9 Lines 64 and Col 10 Lines 1-11 i.e. history map checks each hyperlink URL to determine if it is already listed in the history map, if not the URLs are added and are marked as not being crawled and added to the transaction log. The history map includes a number crawled and number modified data)

In regard to dependent claim 16, in addition to the following reflect similar subject matter claimed in claim 3 and are rejected along the same rationale. (Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction)

As per claims 23 and 25, the applicant discloses the limitations similar to those in claim 21. Claims 23 and 25 are similarly rejected.

5. Claims 4-6 and 17-19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Meyerzon, Blumenthal, and Koike and in further view of Hobbs (US 6523022, filed 7/7/1999).

In regard to dependent claim 4, Meyerzon does not specifically executing one or more Java applets with textual content embedded therein. However, Hobbs mentions that Java applets are used (Hobbs Col 28 Line 35). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Hobbs to Meyerzon, providing Meyerzon the benefit of using Java Applets for web pages in the process of searching the web documents because Java Applets are compatible with many web pages and browsers.

In regard to dependent claim 5, Meyerzon does not specifically mention *wherein the loading secondary documents further comprises the loading of secondary documents including web documents selected from the group of documents consisting of in-line frames, frames, and equivalents*. However, Hobbs mentions that frames and in-line frames are used (Hobbs Col 7 Lines 63 through Col 8 Lines 1-34). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Hobbs to Meyerzon, providing Meyerzon the benefit of using frames and in-line frames for easy viewing for the user.

In regard to dependent claim 6, Meyerzon does not specifically mention *wherein the loading secondary documents further comprises the loading of secondary documents including one or more Java Script components with textual content embedded therein*. However, Hobbs mentions that Java applets are used (Hobbs Col 28 Line 35). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Hobbs to Meyerzon, providing Meyerzon the benefit of using Java Scripts for web pages in the process of searching the web documents because Java Scripts are compatible with many web pages and browsers.

In regard to dependent claim 17, the applicant discloses the limitations substantially similar to those in claim 4 and the same rejection is incorporated herein (Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction).

In regard to dependent claim 18, the applicant discloses the limitations substantially similar to those in claim 5 and the same rejection is incorporated herein

Art Unit: 2178

(Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction).

In regard to dependent claim 19, the applicant discloses the limitations substantially similar to those in claim 6 and the same rejection is incorporated herein (Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction).

6. Claims 12, 22, and 24 are rejected under 35 U.S.C. 103(a) as being unpatentable over Meyerzon, Blumenthal, and Koiske and further in view of Lawrence et al. (US 6289342, filed 20 May 1998, hereafter Lawrence).

In regard to dependent claim 12, Meyerzon discloses *accessing a set of memory structures of the Page Renderer* (Meyerzon Col 6 Lines 23-60 i.e. accessing local and remote memory devices); *copying a text portion of the structures into a text map* (Meyerzon Col 15 Lines 15-16 i.e. copying all of the history map entries into the transaction log as entries); *inspecting any in-line GIF and JPEG image references in the memory structures* (Meyerzon Col 9 Lines 37-46 i.e. an image is retrieved to display on a web page and it is well known in the art the images displayed on web pages can be a gif and jpeg image); *extracting alternate text attributes* (Meyerzon Col 5 Lines 7-8 i.e. extracting data from each of the retrieved documents); *adding the alternate text attributes to a text map* (Meyerzon Col 2 Lines 48-51 i.e. information from the electronic document retrieved from the web crawl is stored in an index); *extracting text content from the GIF and JPEG images; adding text content from the images to the text map*

Art Unit: 2178

(Meyerzon Col 9 Lines 37-46 i.e. an image is retrieved to display on a web page and it is well known in the art the images displayed on web pages can be a gif and jpeg image Col 5 Lines 7-8 i.e. extracting data from each of the retrieved documents Col 2 Lines 48-51 i.e. information from the electronic document retrieved from the web crawl is stored in an index); *and forwarding the text map to a Page Summarizer subroutine.* (Meyerzon Col 9 Lines 64 and Col 10 Lines 1-11 i.e. history map checks each hyperlink URL to determine if it is already listed in the history map, if not the URLs are added and are marked as not being crawled and added to the transaction log. The history map includes a number crawled and number modified data)

Meyerzon does not specifically mention *invoking an optical character recognition engine*; analyzing any in-line GIF and JPEG images *using the optical character recognition engine* for text content. However, Lawrence mentions extracting data using optical character recognition (Lawrence Col 7 Lines 51-56 i.e. conversion to electronic form by use of OCR). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Lawrence to Meyerzon, providing Meyerzon the benefit of extracting content from a document using OCR, which is quicker the typing out an entire document manually by hand.

As per claims 22 and 24, the applicant discloses the limitations substantially similar to those in claim 12. Claims 22 and 24 are similarly rejected.

Response to Arguments

Art Unit: 2178

7. Applicant's arguments with respect to claims 3-13, 16-19, and 21-25 have been considered but are moot in view of the new ground(s) of rejection.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Kyle R. Stork whose telephone number is (571) 272-4130. The examiner can normally be reached on Monday-Friday (8:00-4:30).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Stephen Hong can be reached on (571) 272-4124. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Kyle R Stork
Patent Examiner
Art Unit 2178



STEPHEN HONG
SUPERVISORY PATENT EXAMINER

krs